CURRENT PROTOCOLS IN MOLECULAR BIOLOGY

VOLUME 1

John Wiley & Sons, Inc.

Library of Congress Cataloging in Publication Data:

Current protocols in molecular biology. 4 vols.

1. Molecular biology—Technique. 2. Molecular biology—Laboratory manuals. I. Ausubel, Frederick M.

QH506.C87 1987 574.8'8'028 87-21033
ISBN 0-471-50338-X

Printed in the United States of America

20 19 18 17 16

VECTORS DERIVED FROM PLASMIDS

Introduction to Plasmid Biology

Plasmids are self-replicating, extrachromosomal DNA molecules found in virtually all bacterial species. In nature, plasmids occur in exuberant profusion, varying in structure, size, mode of replication, number of copies per bacterial cell, ability to propagate in different bacteria, transferability between bacterial species, and perhaps most important, in the traits they carry. Most prokaryotic plasmids are double-stranded circular DNA molecules; however, linear plasmids have been identified in both gram-positive and gram-negative bacteria. The size of plasmids varies widely, from several kilobases to hundreds of kilobases. Replication of plasmids depends on host-cell proteins but also may require plasmid-encoded functions. Plasmid replication may be synchronized with the bacterial cell cycle, resulting in a low number of plasmid molecules per bacterial cell, or independent of the host cell cycle, allowing for the proliferation of hundreds of plasmid copies per cell. Some plasmids freely transfer their DNA across bacterial species, some only transfer their DNA into bacteria of the same species, and some do not transfer their DNA at all. Plasmids carry genes that specify a wide variety of functions including: resistance to antibiotics, resistance to heavy metals, sensitivity to mutagens, sensitivity or resistance to bacteriophages, production of restriction enzymes, production of rare amino acids, production of toxins, determination of virulence, catabolism of complicated organic molecules, ability to form symbiotic relationships, and ability to transfer DNA across kingdoms.

Starting in the 1970s, vectors for propagation, manipulation, and delivery of specific DNA sequences were constructed with fragments from naturally occurring plasmids, primarily *Escherichia coli* plasmids. All plasmid vectors contain three common features: a ***replicator***, a ***selectable marker***, and a ***cloning site***. The replicator is a stretch of DNA that contains the site at which DNA replication begins (the origin of replication or *ori*), and that also includes genes encoding whatever plasmid-encoded RNAs and proteins are necessary for replication. The selectable marker, necessary for following and maintaining the presence of the plasmid in cells, is usually dominant and is usually a gene encoding resistance to some antibiotic. The cloning site is a restriction endonuclease cleavage site into which foreign DNA can be inserted without interfering with the plasmid's ability to replicate or to confer the selectable phenotype on its host. Over the years, plasmid vectors have increased in sophistication; in addition to these three basic elements many vectors now include features that make them particularly suitable for specific types of experiments.

UNITS 1.6 & 1.7 describe procedures for making plasmid DNA. The process by which plasmids are introduced into *E. coli* is called ***transformation***. Transformation protocols are given in *UNIT 1.8*. For a list of vectors and their salient features see *APPENDIX 5*.

REPLICATORS

One of the ways that replicators are classified is based on the number of plasmid molecules maintained per bacterial cell under some set of standard growth conditions, the so-called ***copy number*** of the plasmid. This book defines ***high-copy-number plasmids*** as those which exist in ≥20 copies per bacterial cell grown in liquid LB medium, and ***low-copy-number plasmids*** as those which exist in <20 copies per cell. High-copy-number-plasmids tend to be smaller than low-copy-number plasmids. They are more commonly used in molecular biological techniques since it is easier to prepare large quantities of pure plasmid DNA from cells that bear them. Low-copy-number plasmids are utilized when it is important to control the gene dosage of a cloned sequence, for example when the DNA sequence or the protein it encodes makes the host organism sick (see Table 1.5.1).

High-copy-number plasmids tend to be under *relaxed control* of replication. These relaxed plasmids initiate DNA replication in a process controlled by plasmid-encoded functions (see Mechanism of Replication and Copy-Number Control), and replication does not depend on the unstable host replication initiation proteins synthesized at the start of the bacterial cell cycle. Because their replication depends only on the stable host enzymatic replication machinery, the copy number of these plasmids can be greatly increased, or amplified, by treatment of the plasmid-containing cells with protein synthesis inhibitors such as chloramphenicol

Contributed by Rhonda Feinbaum
Current Protocols in Molecular Biology (1998) 1.5.1-1.5.17

Table 1.5.1. Characteristics of Commonly Used Plasmid Replicators

Replicator	Prototype plasmid	Size (bp)	Markers on prototype	Copy[a] number	References
pMB1	pBR322	4,362	Ap^r, Tet^r	High; 100-300	Bolivar et al., 1977
ColE1	pMK16	~4,500	Kan^r, Tet^r, $ColE1^{imm}$	High; >15	Kahn et al., 1979
p15A	pACYC184	~4,000	Eml^r, Tet^r	High; ~15	Chang et al., 1978
pSC101	pLG338	~7,300	Kan^r, Tet^r	Low; ~6	Stoker et al., 1982
F	pDF41	~12,800	TrpE	Low; 1 to 2	Kahn et al., 1979
R6K	pRK353	~11,100	TrpE	Low; <15	Kahn et al., 1979
R1 (R1*drd-17*)	pBEU50	~10,000	Ap^r, Tet^r	Low at 30°C; high above 35°C[b]	Uhlin et al., 1983
RK2	pRK2501	~11,100	Kan^r, Tet^r	Low; 2 to 4	Kahn et al., 1979
λ *dv*	λ *dvgal*	—[c]	Gal	—	Jackson et al., 1972

[a]Copy numbers are for the prototype plasmid. Plasmid vectors that contain replicators derived from these plasmids may have different copy numbers due to introduction of mutations into the replicator. For example, pUC series (pmB1-derived) has copy numbers of 1000-3000.

[b]Temperature sensitive.

[c]Not known.

col or spectinomycin. These protein synthesis inhibitors prevent synthesis of the replication-initiation proteins required for chromosomal replication, allowing the replication enzymes to be commandeered for plasmid replication. High-copy-number plasmids usually do not have any mechanism to ensure correct segregation of the plasmids to the daughter cells; they rely on random assortment to partition at least one copy of the plasmid to each daughter.

Low-copy-number plasmids are usually under *stringent control*. Initiation of replication of these plasmids depends on unstable proteins synthesized at the start of the bacterial cell cycle and thus is synchronized with the replication of the bacterial chromosome. As copy number decreases, random segregation of plasmid copies is not sufficient to ensure that each daughter cell acquires a copy of the plasmid. However, most low-copy-number plasmids carry genes that guarantee their maintenance in the bacterial population. Stabilizing loci that encode a mechanism for postsegregational killing of plasmid-free daughter cells and a system for efficiently resolving plasmid multimers so that they can be appropriately partitioned have been identified. In addition, *cis*- and *trans*-acting partitioning loci (*par*) that are thought to constitute an active mechanism for distribution of plasmid copies to daughter cells have also been identified.

MECHANISM OF REPLICATION AND COPY-NUMBER CONTROL

While there are many different replicators, the majority of plasmid vectors used in routine recombinant DNA work contain one of the functionally similar replicators derived from plasmids ColE1 or pMB1 (for example, the popular pBluescript series contain a ColE1 origin, and the pUC plasmids are derived from pMB1). The ColE1 replicator is a 600-nucleotide DNA fragment that contains the origin of replication (*ori*), encodes an RNA primer, and encodes two negative regulators of replication initiation. All enzymatic functions for replication of the plasmid are provided by the bacterial host.

Plasmid replication begins with transcription of an RNA primer upstream of the *ori* (RNAII; see Fig. 1.5.1) by the host RNA polymerase. RNAII is elongated through and terminated downstream of the *ori*. Interaction of a specific secondary structure in the nascent RNAII transcript with the DNA template results in formation of a persistent hybrid between RNAII and the DNA template such that RNAII remains paired with the DNA template at the *ori*. The RNAII transcript is cleaved by RNAseH at the *ori* sequence. This processed RNAII primer is extended by DNA polymerase I to initiate plasmid replication. Regulation of the proper RNAII secondary structure controls initiation of DNA replication and is responsible

for determining the number of plasmid molecules per cell.

Both ColE1 and pMB1 plasmids are high-copy-number plasmids, maintained at between 15 and 25 copies per bacterial cell, respectively. The copy number of these plasmids is regulated by an antisense RNA transcript, RNAI, and the protein ROP, the product of the *rop* gene. RNAI and the ROP protein act in concert to intercept formation of the proper RNAII secondary structure. RNAI is exactly complementary to the 5′ end of the RNAII transcript. RNAI base pairs with the 5′ end of RNAII, preventing the formation of the specific secondary structure necessary for establishment of a persistent hybrid between RNAII and the DNA template that is a prerequisite for maturation of the RNAII primer. ROP protein stabilizes the interaction between the antisense RNAI and the primer RNAII. Mutations that disrupt or destabilize

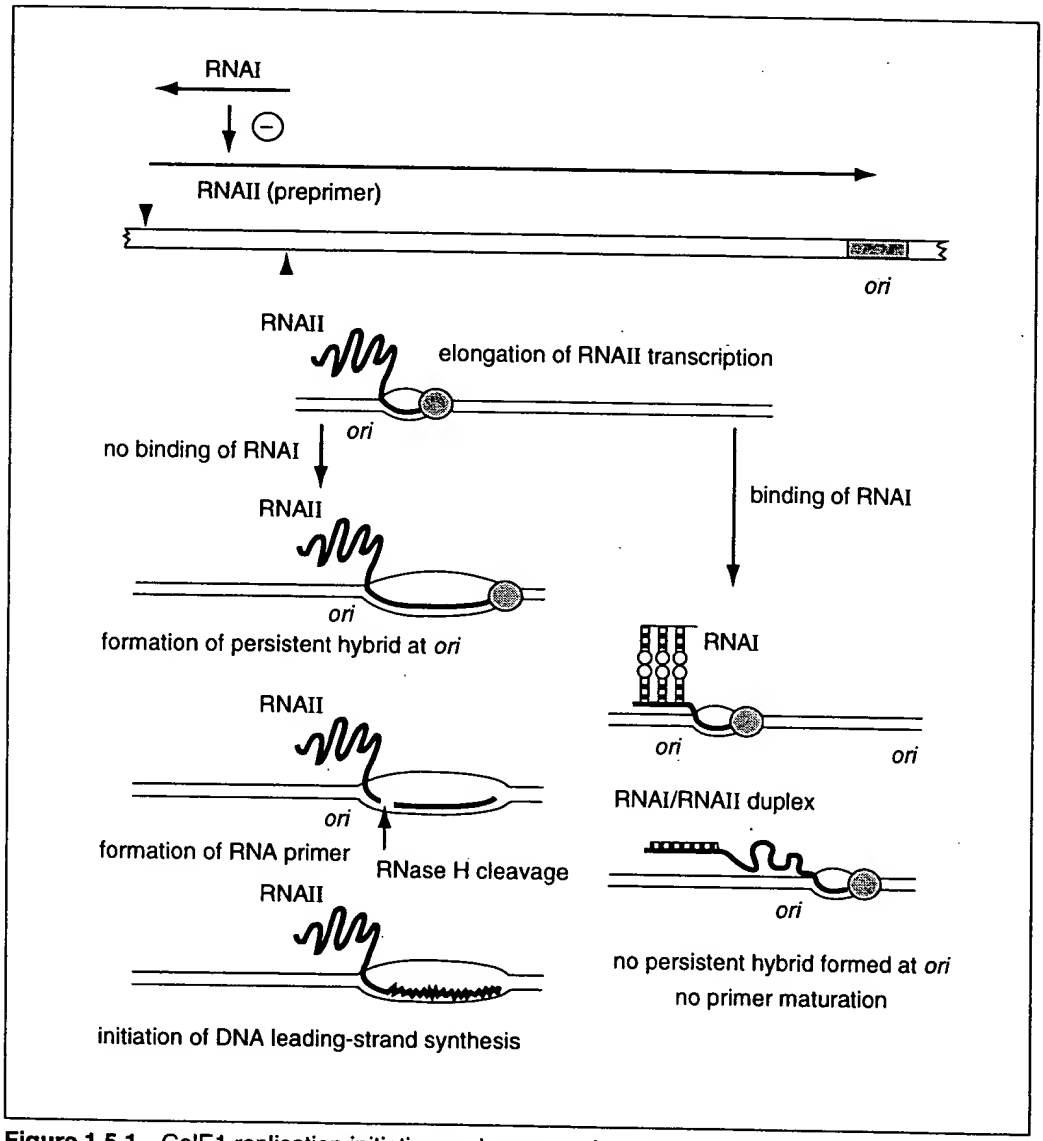

Figure 1.5.1 ColE1 replication initiation and copy number control. ColE1 replication requires the formation of a RNA primer to initiate DNA synthesis. The RNA primer is derived from processing of a transcript, RNAII (bold), that starts upstream of the *ori*. Appropriate processing of the RNAII transcript is dependent on formation of a persistent hybrid between RNAII and the *ori* DNA template. If the proper RNAII secondary structure forms, then the RNA/DNA duplex is maintained at the *ori* and RNAse H can cleave the RNAII transcript to generate the primer for DNA synthesis. Processing of RNAII to form the primer is regulated by a second transcript, RNAI. RNAI is complementary to the 5′ end of RNAII. If the RNAI transcript forms a duplex with RNAII, RNAII cannot take on the secondary structure necessary to create the persistent hybrid at the *ori* and maturation of the RNAII primer is inhibited. The copy number of ColE1 plasmids is determined by the balance between successful RNAII processing events and those inhibited by RNAI. Adapted from Gerhart et al. (1994) with permission from the *Annual Review of Microbiology*.

the RNAI/RNAII interaction result in a higher plasmid copy number. For example, the pUC plasmid series have pMB1 replicators that do not contain an intact *rop* gene, which accounts for a 2-fold increase in copy number over plasmids with intact pMB1 replicons. Interestingly, additional mutation(s) in the origin are the main factor in the ultra-high-copy number (1000 to 3000) of this plasmid series (K. Struhl, unpub. observ.)

PLASMID INCOMPATIBILITY

For experiments that require that more than one plasmid vector be maintained in a bacterial cell at the same time, another critical feature of plasmid replicators is whether or not two plasmid replicators are *compatible*. Two plasmids are said to be *incompatible* with one another, and hence belong to the same *incompatibility group*, if they cannot stably co-exist. Plasmids are generally incompatible if they share any function required for the regulation of plasmid replication. For example, ColE1- and pMB1-derived plasmids are incompatible with one another but are compatible with p15A plasmids. The incompatibility of ColE1 and pMB1 plasmids is a consequence of two facts: first, that plasmid DNA replication of these plasmids is negatively controlled by RNAI which acts in *trans* on other plasmids with the same primer RNA, and second, that these plasmids lack a mechanism to ensure that each plasmid in a cell replicates once per cell cycle. Therefore, if a cell that contains a pMBI-derived plasmid is subsequently transformed with a ColE1-derived plasmid, cells selected to contain the ColE1 plasmid will usually have lost the pMBI plasmid.

SELECTABLE MARKERS

In order to guarantee that a plasmid vector is taken up by and maintained in bacterial cells, there must be a way to select for plasmid-containing cells. Genes encoding resistance to antibiotics such as ampicillin, tetracycline, kanamycin and chloramphenicol are the most common bacterial selectable markers for plasmid vectors. Typically, cells are transformed with plasmid DNA using the technique described in *UNIT 1.8* and then plated out on LB plates that contain the proper antibiotic (see recipes in *UNIT 1.1*). Only the bacterial cells containing the plasmid will grow on the selective medium; the antibiotic-resistance phenotype conferred is dominant to the antibiotic-sensitive phenotype of cells that do not possess the plasmid vector. Another dominant selectable marker that is occasionally used is the immunity to infection by phage lambda (lambda repressor). In addition, recessive markers are sometimes used in plasmid selections; for example, *leuB*$^-$ *E. coli* cannot grow in the absence of leucine, and selection for growth of these strains in the absence of leucine allows isolation of colonies transformed with a plasmid that contains a gene that complements *leuB*.

For experiments that involve introduction of plasmid vectors into systems other than bacteria, such as yeast or mammalian cells, it is usually necessary to select for the presence of plasmid DNA in these hosts. Selectable markers for yeast and mammalian systems are generally distinct from those used in *E. coli* (see Plasmid Vectors for Yeast and Plasmid Vectors for Expression in Cultured Mammalian Cells). In cases where a vector is needed that will be used in both *E. coli* and another host, it is necessary for the plasmid vector to carry selectable markers for both hosts.

CLONING SITE

Today's plasmid vectors contain a multiple cloning site (MCS) or ***polylinker*** cloning region that can include >20 tandemly arranged restriction endonuclease sites. The sites in the polylinker are almost always designed to be unique within the vector sequence so that cutting the vector with a restriction endonuclease in the polylinker and cloning foreign DNA into this site does not disrupt other critical features of the vector. The plethora of sites in the polylinker ensures that the appropriate enzyme sites will be available for cloning most DNA fragments, provides unique reference restriction sites for rapid restriction mapping of the insert, and generally allows for a great deal of flexibility when manipulating the cloned DNA.

The sequences that directly flank the polylinker site are often useful for manipulation or analysis of insert DNA. Many polylinker sites are flanked by sequences for which there are commercially available complementary oligonucleotides, for example the M13 reverse, –20, and –40 primers, that can be used for priming polymerase chain reactions (PCR) or DNA sequencing reactions. Such primers are useful tools for amplification or sequencing of any DNA fragment inserted into the polylinker. Some polylinkers are bordered by 8-bp-cutter restriction sites, like *Not*1. These sites occur infrequently in DNA and thus allow for the easy excision of an intact insert fragment from the plasmid vector.

Many plasmid vectors have been developed with bacteriophage—SP6, T7, or T3—promoters flanking the polylinker cloning sites. These promoters can be used in vitro or in vivo, if the bacteriophage RNA polymerase is also present, to produce large quantities of RNA transcripts from DNA inserted into the polylinker.

In order to make it easier to identify plasmids that contain insert DNA, the polylinkers of some vectors have been engineered so that introduction of DNA into the polylinker results in a scorable phenotype. The most common example of this is insertion of the polylinker of many basic cloning vectors, the pUC series for example, into the *lacZ* gene α fragment. In the appropriate genetic background, production of the lacZ α fragment allows for formation of an active β-galactosidase enzyme which results in the formation of blue colonies on Xgal/IPTG indicator plates. Cloning into these polylinkers prevents production of a functional lacZ α fragment, allowing for rapid identification of plasmids containing inserts as white colonies on Xgal/IPTG indicator plates. Vectors have also been developed that allow for direct selection of plasmids containing inserts by location of the polylinker in the *ccdB* (control of cell death) gene which causes cell death in *E. coli*. Disruption of the *ccdB* gene by introduction of an insert into the polylinker allows the cells to survive, and thus only recombinants will grow under conditions where the *ccdB* gene is expressed.

CHOOSING A PLASMID VECTOR

Some of the basic features to consider when selecting a plasmid vector include the size of the vector, its copy number, the polylinker, and the ability to select and/or screen for inserts. Large plasmids, >15 kb, do not transform well and frequently give lower DNA yields. Consider the final size of the vector plus insert when planning an experiment and wherever possible use smaller vectors. The higher the copy number the more vector DNA is produced, but high-copy-number vectors may not be applicable to all situations (see Mechanism of Replication and Copy-Number Control). When choosing a plasmid vector, consider both what sites are present in the polylinker and the order of the sites. If it will be necessary to manipulate the cloned sequence subsequent to insertion into the polylinker, plan ahead to ensure that the necessary sites will remain available in the polylinker. Selections or screens for identification of recombinant clones are useful for experiments where cloning efficiency is expected to be low or when generation of a large number of clones is necessary. However, for routine subcloning experiments the advent of PCR (*UNIT 15.1*) has made it possible to rapidly screen through a large number of transformants to identify potential recombinant molecules, obviating the need for histochemical and genetic screening methods.

The primary factor in choosing a plasmid vector is to understand and anticipate the experiments for which the recombinant clone will be used. The specialized functions of plasmid vectors are generally the key to selecting the correct vector for an experiment. For example, completely different types of vectors would be used for generating large quantities of DNA, expressing a fusion protein in bacteria, or for undertaking a two-hybrid screen in yeast. Once the type of vector required is determined then deciding upon a particular vector is dependent on both the details of the ancillary vector features, for example the type of promoter used to express a recombinant protein, and the properties of the replicator, polylinker, and selectable marker.

PLASMID VECTORS FOR PRODUCTION OF SINGLE-STRANDED DNA

Plasmids have been developed that contain a filamentous phage origin of replication in addition to a plasmid *ori*. These "phagemid" vectors (*UNIT 1.14*) can be grown and propagated as plasmids. However, upon super-infection of a plasmid-containing cell with a wild-type helper phage, the phage *ori* becomes active, and single-stranded DNA (ssDNA) is produced and secreted. There are usually (+) and (−) versions of these vectors where the phage *ori* is in opposite orientation so that it is possible to produce ssDNA from either DNA strand. For many years, ssDNA was the substrate of choice for DNA sequencing (*UNIT 7.4*) and oligonucleotide-directed mutagenesis (*UNIT 8.1*). Development of sequencing and mutagenic protocols that use double-stranded templates has made the production of ssDNA a less frequently utilized feature of plasmid vectors. The pBluescriptI, pBluescriptII, and pBS phagemid vectors derived from the general cloning vector pUC19 (Fig. 1.5.2) are examples of phagemids that incorporate the f1 filamentous phage *ori*.

PLASMID VECTORS FOR CLONING LARGE INSERTS

Cosmid vectors, plasmids carrying a lambda phage *cos* site (e.g., pWE15, Fig. 1.5.3), were

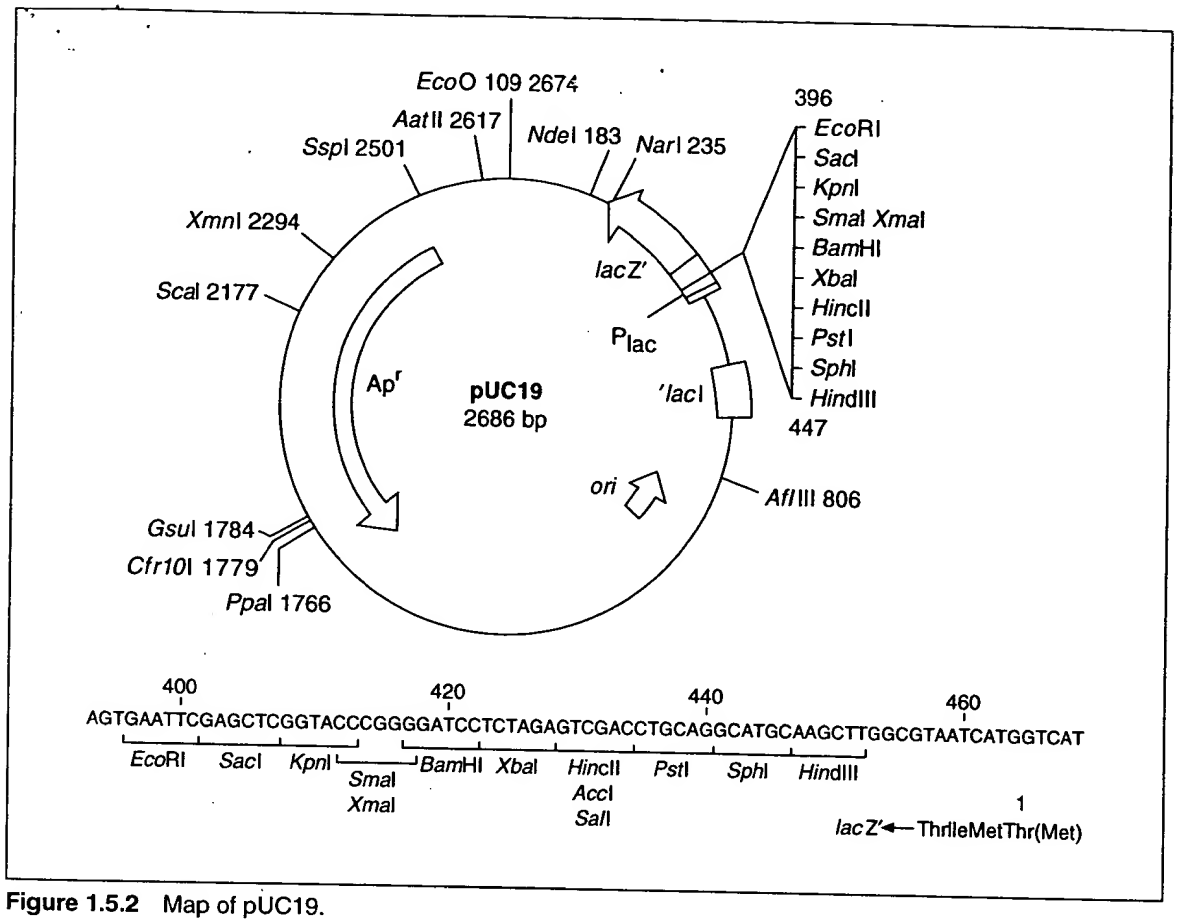

Figure 1.5.2 Map of pUC19.

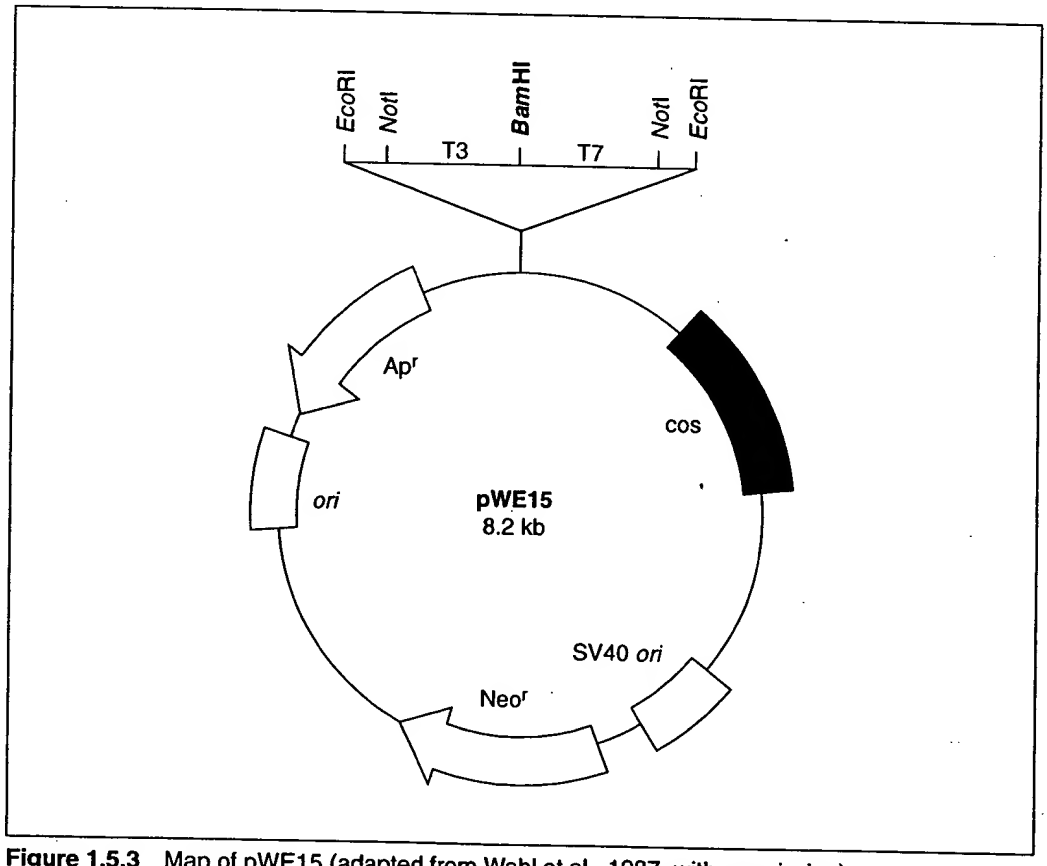

Figure 1.5.3 Map of pWE15 (adapted from Wahl et al., 1987, with permission).

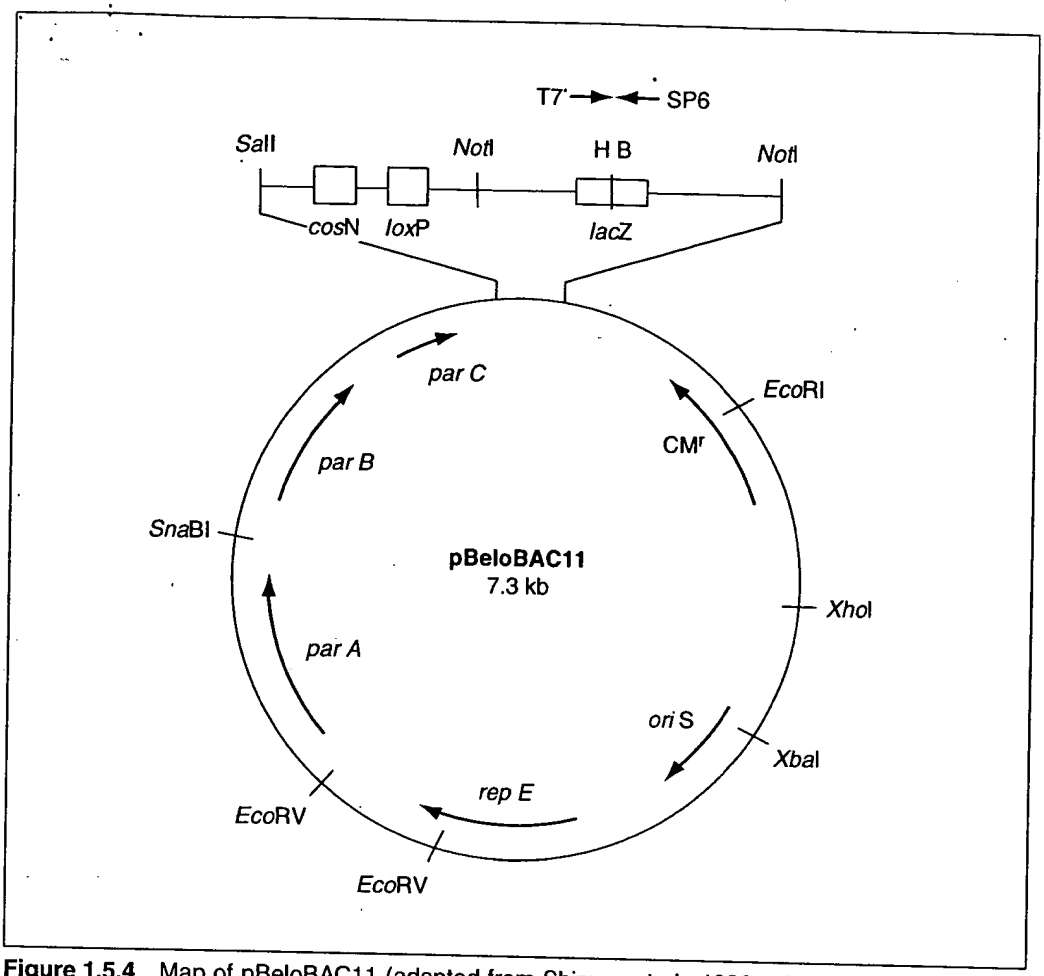

Figure 1.5.4 Map of pBeloBAC11 (adapted from Shizuya et al., 1992, with permission). Abbreviation: CM, chloramphenicol.

developed to facilitate cloning of large DNA fragments (*UNIT 1.10*). Cosmids can be transformed into cells like plasmids and once in the cells, replicate using their plasmid *ori*. The ColE1 type replicators are the most commonly used in these vectors, and cosmids can generally be maintained at high-copy-number in *E. coli*. Cosmids can also be packaged into lambda phage heads. In order for packaging to occur the *cos* sites must be separated by 40 to 50 kb, the approximate length of the wild-type lambda genome. Many cosmid vectors are between 5 to 10 kb in size and therefore can accept inserts of 30 to 45 kb. Because of the size of inserts they can accept, the high efficiency of packaging recombinant molecules into phage, and the efficiency of infection of cosmid-containing phage heads, cosmid vectors are frequently used for making genomic libraries. Unfortunately, propagation of insert-containing cosmid vectors in *E. coli* sometimes results in the deletion of all or a portion of the insert. To address this problem, a new set of cosmid vectors have been developed that replace the ColE1 replicator with the F factor replicator. These "fosmid" vectors are maintained at low-copy-number, 1 to 2 copies per cell, and are more stable than higher-copy-number cosmid vectors when grown in *E. coli* (Kim et al., 1992).

Another type of plasmid cloning vector, called bacterial artificial chromosome (BAC; Fig. 1.5.4), has been developed using the F factor replicator for propagation of very large pieces of DNA (100 to 500 kb). The vectors are used in a similar manner to yeast artificial chromosome (YAC) vectors but have the advantage of being manipulated solely in *E. coli*.

PLASMID VECTORS FOR EXPRESSION OF LARGE QUANTITIES OF RECOMBINANT PROTEINS

There are a wide variety of vectors for expressing high levels of recombinant proteins in *E. coli*, insect, and mammalian cells (Chapter 16). The general goal of expressing proteins in any of these systems is to produce large quantities of a particular protein upon demand.

While the features of the protein expression systems vary considerably, the basic properties outlined for expressing proteins in *E. coli* are common to all of them. Expression vectors are usually designed such that production of the foreign protein is tightly regulated. This is necessary because the host cellular machinery is co-opted to produce large quantities of the foreign protein, the shear amount of which may be toxic to the cell, and/or the foreign protein may encode a function that will inhibit cell growth or kill the host cell. Generally, expression vectors are configured so the polylinker cloning site is downstream of an inducible promoter. One of the promoters commonly used in *E. coli* expression vectors is the hybrid trp/lac promoter (trc) which contains the lacO operator site. This promoter is turned off in the presence of the $lacI^q$ repressor; the repressor gene is either carried by the bacterial host or is encoded on the expression vector itself. Expression of the foreign protein from the trc promoter is induced by the addition of IPTG.

The quantity of foreign protein produced will be determined by both the rate of transcription of the gene and the efficiency of translation of the mRNA. Therefore, in addition to regulated highly inducible promoters, many *E. coli* protein expression vectors are designed to optimize translation of the foreign protein in bacterial cells. These expression vectors include an efficient ribosome binding site and an ATG start codon uptsream of the polylinker cloning site. Usually the cloning sites in the polylinker are designed so that it is possible to make an in-frame fusion to the protein of interest in all three reading frames.

For many experiments, high levels of pure recombinant protein are required, and some expression vectors are designed to create tagged or fusion proteins that facilitate purification of the recombinant protein. Tag sequences may be located 5′ or 3′ to the polylinker, creating either amino- or carboxy-terminal-tagged fusion proteins. Six polyhistidine residues (*UNIT 10.11B*), the FLAG epitope, and glutathione-*S*-transferase protein (*UNIT 16.7*) are some of the sequences that are appended to proteins to assist in purification. Tagged or fusion proteins can be rapidly and efficiently purified using an appropriate affinity column designed to tightly bind the tag or fusion region. Many protein expression vectors are created with specific protein cleavage sites adjacent to the tag or fusion sequences to allow for removal of these sequences from the purified protein. This feature may prove to be essential if the tag or fusion sequence impairs the function of the protein in the relevant assays.

PLASMID VECTORS FOR REPORTER GENE FUSIONS

Plasmid vectors have been designed to simplify the construction and manipulation of reporter gene fusions, where a promoter of interest is used to drive an easily scored marker gene (*UNIT 9.6*). Gene fusions provide a rapid and simple method for following the expression pattern conferred by a particular promoter. There are generally two types of reporter fusions, transcriptional and translational fusions. For transcriptional fusions, the ATG start codon is provided by the marker gene. In translational fusions, the 5′ untranslated region and ATG are provided by the gene of interest; in fact these constructs may fuse a large portion, or even entire coding region, to the amino terminus of a marker. In the reporter vector, the polylinker cloning site is located directly upstream of the reporter gene for insertion of the promoter fragment. In vectors that are designed for expression in eukaryotic cells, a polyadenylation signal is located downstream of the reporter gene. There are a variety of reporter genes used including chloramphenicol acetyltransferase, luciferase, β-galactosidase, secreted alkaline phosphatase, human growth hormone, β-glucuronidase, and green fluorescent protein (also see *UNITS 9.6-9.7C*). An example of a reporter vector pEGFP, is shown in Figure 1.5.5. The choice of reporter gene will depend on the cell type or organism, whether the assay will be done in vivo or in vitro, and whether quantitative or qualitative data is desired. The replicator, selectable marker, and other elements of the reporter vector will also have to be compatible with the system.

PLASMID VECTORS FOR YEAST

Yeast plasmid vectors contain the same basic features as *E. coli* vectors—replicator, selectable marker, and cloning site (*UNIT 13.4*). There are two primary types of replicators used in yeast plasmid vectors, autonomously replicating sequences (ARS) derived from the yeast chromosomes and the natural yeast 2μm plasmid replicator. ARS-containing plasmids frequently contain yeast centromeric sequences to ensure their stable maintenance in the population of cells. Both ARS and 2μm vectors average 10 to 30 copies per cell. Many yeast vectors are "shuttle" vectors that can be maintained in both *S. cerevisae* and *E. coli* (e.g., pRS303; see Fig. 1.5.6). These vectors have an *E. coli* plas-

mid replicator, frequently pMB1 derived, and a yeast replicator. Conversely, the yeast integrating plasmid vectors, used for introduction of genes into the yeast chromosome, have a bacterial replicator but no yeast replicator.

The selectable markers in yeast are for the most part recessive markers, usually cloned yeast genes that are used to complement mutations in a biosynthetic pathway. For example, the *URA3* marker carried on a plasmid is used to restore the ability to grow in the absence of uracil to a *ura3*$^-$ yeast. Therefore, yeast selectable markers, unlike most bacterial selectable markers, are strain dependent.

There are a wide variety of yeast vectors available that have a range of specialized func-

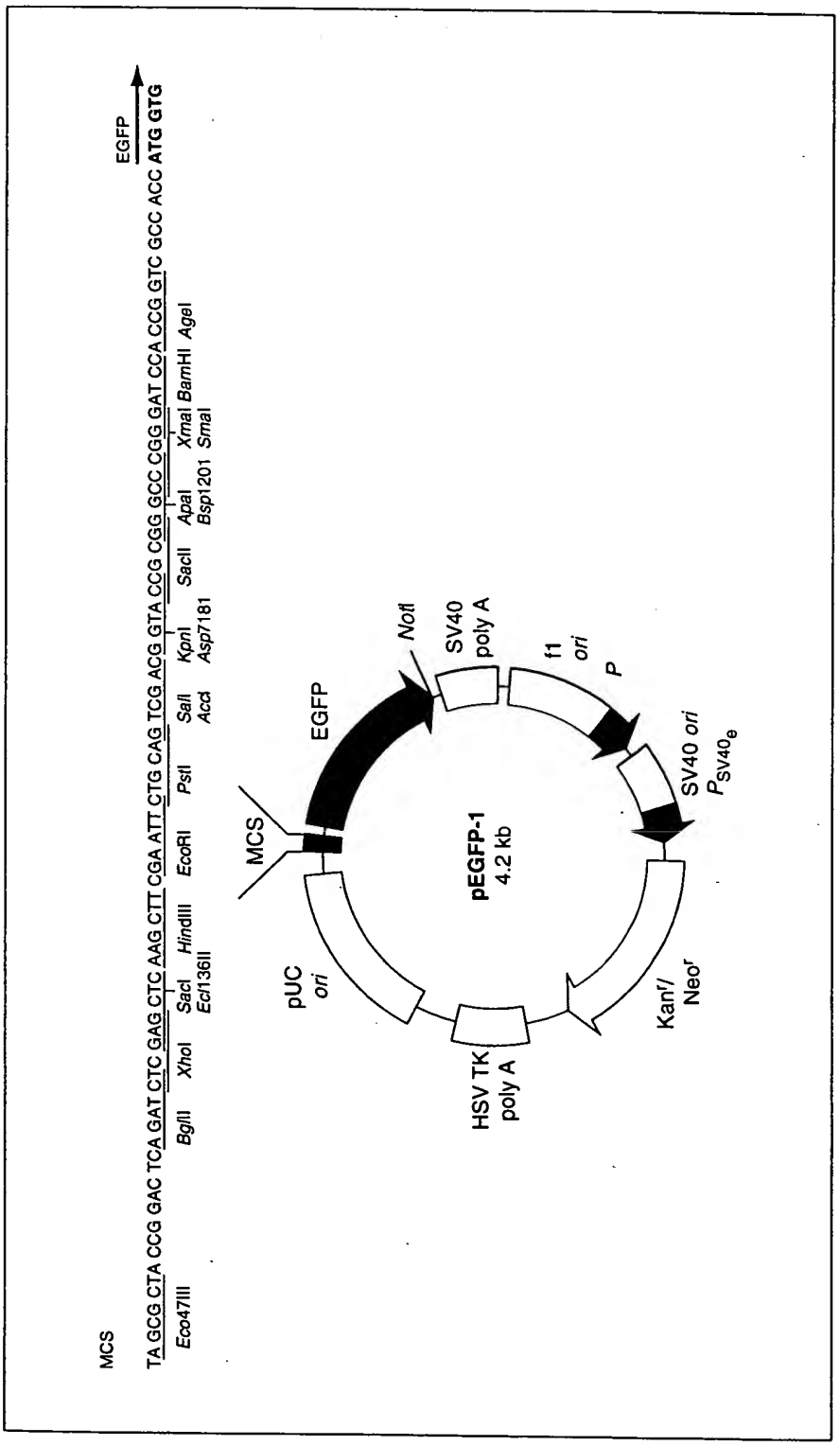

Figure 1.5.5 Map of pEGFP-1.

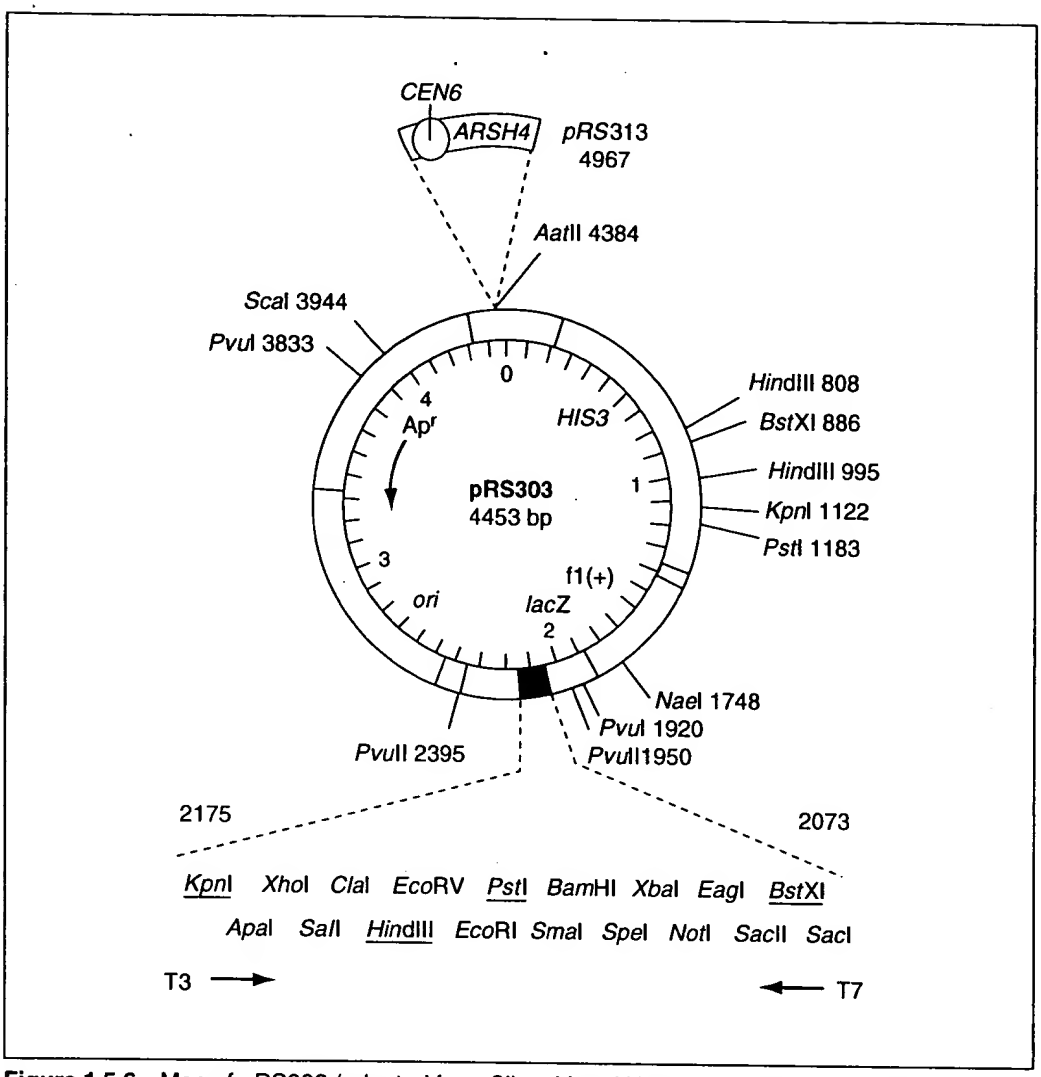

Figure 1.5.6 Map of pRS303 (adapted from Sikorski and Heiter, 1989, with permission).

tions—e.g., expression of recombinant proteins in yeast, integration of sequences into the yeast genome, and cloning of very large fragments (hundreds of kilobases) of genomic DNA.

PLASMID VECTORS FOR EXPRESSION IN CULTURED MAMMALIAN CELLS

The type of vector that is used in mammalian cells depends on whether the experiment involves transient transfection into mammalian cells or the generation of stable mammalian cell lines carrying the construct of interest (*UNITS 9.1-9.5*). Virtually any plasmid vector that contains an appropriate construct for expression in mammalian cells can be used in transient assays. In transient assays, the plasmid vector carrying the DNA of interest is transfected into mammalian cells, the cells are harvested some time later (24 to 96 hr), and the pertinent assay is performed. It is not necessary for plasmid vectors used in these assays to carry a mammalian selectable marker or to replicate in mammalian cells; therefore easily manipulatable bacterial plasmids, like pUC, are usually the vectors of choice.

Plasmid vectors carrying a selectable marker that functions in mammalian cells are necessary for the generation of stable transgenic lines (*UNITS 9.5*; e.g., pcDNA3.1, Fig. 1.5.7). In order to generate a stable mammalian cell line, the plasmid DNA is transfected into the mammalian cells, and over a period of several weeks the DNA of interest is selected for based on expression of the vector-borne marker. Many mammalian vectors cannot replicate in mammalian cells, and the only way to maintain the DNA of interest and the selectable marker is for the vector to randomly integrate into the mammalian genome. However, there are some plasmid vectors that carry the simian

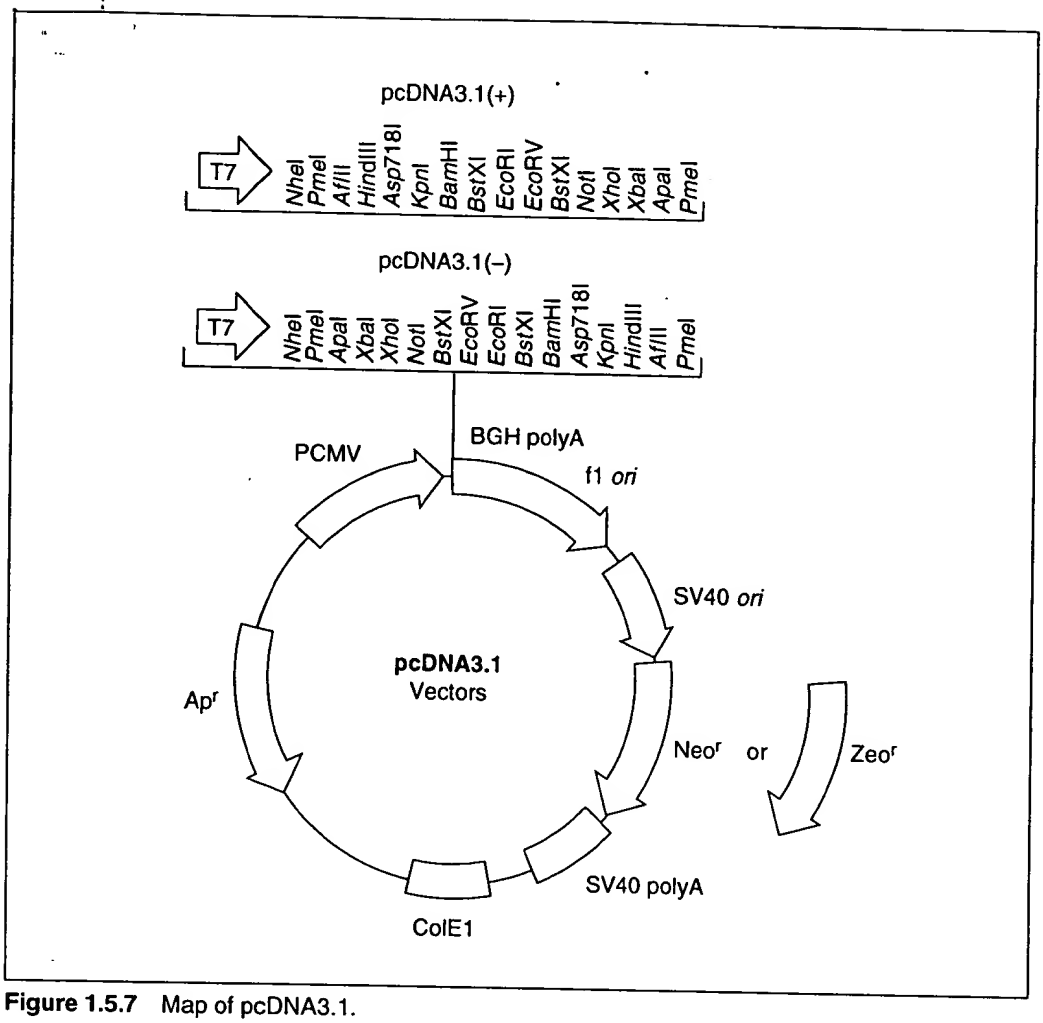

Figure 1.5.7 Map of pcDNA3.1.

virus (SV40) or bovine papilloma virus (BPV) *ori* that can replicate in mammalian cells if the necessary viral replication proteins are provided; in most cases the viral replication proteins must be provided by the host cell line, limiting the range of cell types in which these vectors are useful.

The most critical general feature of plasmid vectors (and also viral vectors) for creation of stable mammalian cell lines is the selectable marker (see *UNIT 9.5*). Unlike transient assays, formation and maintenance of stable cell lines requires selection. The number of selectable markers available for mammalian cells is limited; resistance to hygromycin, puromycin, G418, and neomycin are the predominant markers used. Since expression of foreign proteins or assaying the expression of a mammalian gene may require multiple plasmids and/or integration of constructs into the mammalian chromosome, careful planning must take place to ensure that all constructs can be selected for with the limited number markers. Further, since selection may be necessary over a long period of time—weeks for generation of the lines and years for their maintenance—it is important to recognize that some of the antibiotics used in selection are very expensive and thus cost may be a factor in experimental design.

Specialized plasmid vectors are also used for production of viruses that can infect mammalian cells. Plasmid vectors have been designed to produce infectious retroviral particles when transfected into the appropriate packaging cell line (*UNIT 9.9*). These retroviral vectors are then used to create stable transgenic lines in mammalian cell types not amenable to transfection. In addition, plasmid vectors have been designed to allow easy insertion of DNA sequences into vaccinia virus for the purpose of creating recombinant viruses that overexpress recombinant proteins (*UNIT 16.15*).

PLASMID VECTORS FOR NON-*E. COLI* BACTERIA

Three features required of all bacterial plasmid vectors are that they replicate (unless they are suicide vectors), carry a selectable marker,

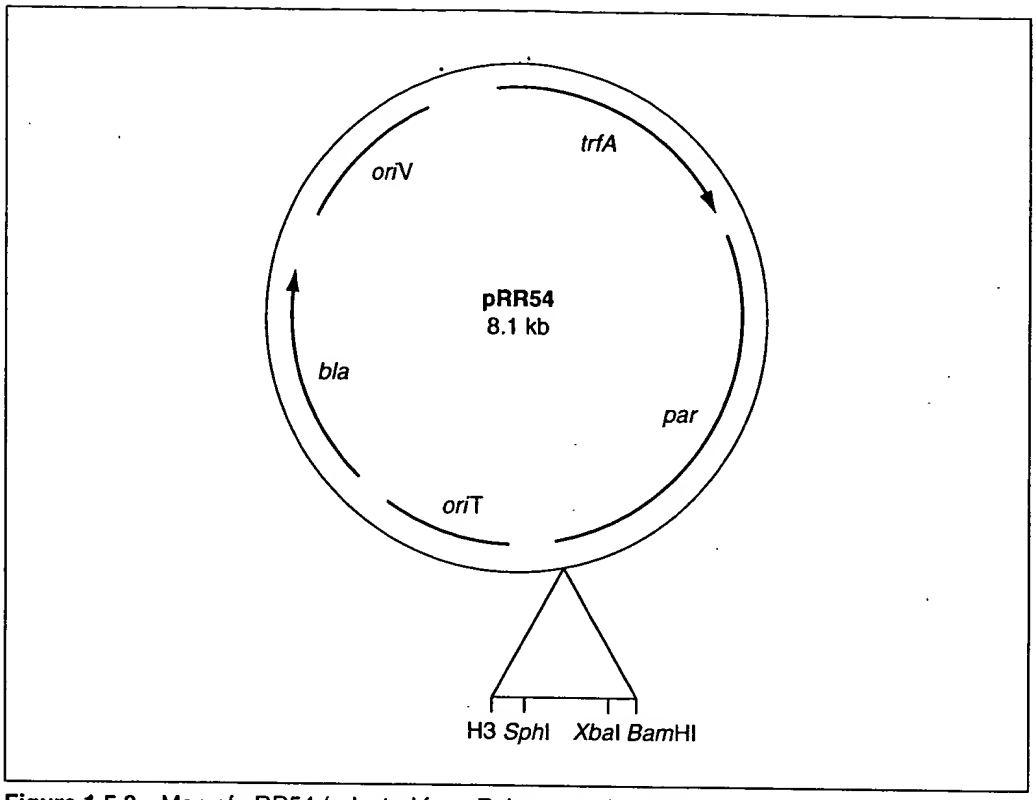

Figure 1.5.8 Map of pRR54 (adapted from Roberts et al., 1990, with permission).

and can be easily introduced into host cells. The first thing to consider when selecting a plasmid vector for use in a non-*E. coli* host is whether or not it can replicate and be stably maintained in the particular strain. Different plasmid replicators have different host ranges, some have a narrow host range and can only replicate in a specific strain while others are promiscuous and can replicate in a wide variety of host (e.g., pRR54, Fig. 1.5.8). Unfortunately, the ColE1-type replicators have a narrow host range, thus many standard *E. coli* vectors cannot be maintained in other bacteria. However, a number of broad-host-range replicators, such as RK2 and RSF1010, have been well characterized and used to construct vectors that can replicate in many gram-negative bacterial species (and in the case of RSF1010, some gram-positive species as well).

Antibiotic resistance genes are used as selectable markers in non-*E. coli* bacterial hosts; however, the quantity of an antibiotic used to select against non-plasmid-containing cells is usually higher than the quantity used for *E. coli* selection. Furthermore, some bacterial strains are inherently resistant to particular antibiotics, thus it is important to determine whether the selectable marker carried on a plasmid vector is functional in a particular strain.

The favored method for introduction of plasmid DNA into bacterial host cells varies widely with the bacterial strain. Unlike *E. coli*, many bacteria cannot be efficiently transformed by chemical procedures or electroporation. In these cases bacterial mating is used to introduce plasmid DNA into the desired bacterial host. Mating to transfer a plasmid vector from an *E. coli* host where the vector is maintained and manipulated to a recipient bacterium requires both *cis*- and *trans*-acting functions. The *tra* (or *mob*) genes encode the *trans*-acting proteins necessary to transfer the plasmid DNA from one bacterium to another, and they are usually located on a helper plasmid that is distinct from the plasmid vector. Any plasmid vector that is to be mobilized must contain the *cis*-acting site called *oriT* where the DNA is cleaved and transfer is initiated.

MAPS OF PLASMIDS

Figures 1.5.2-1.5.11 present maps of plasmids that are in widespread use or are examples of plasmids whose special functions make them useful for particular techniques described in this manual. Note that the trend in development of vectors is to include multiple features on a single vector, and many of these examples span the vector categories described.

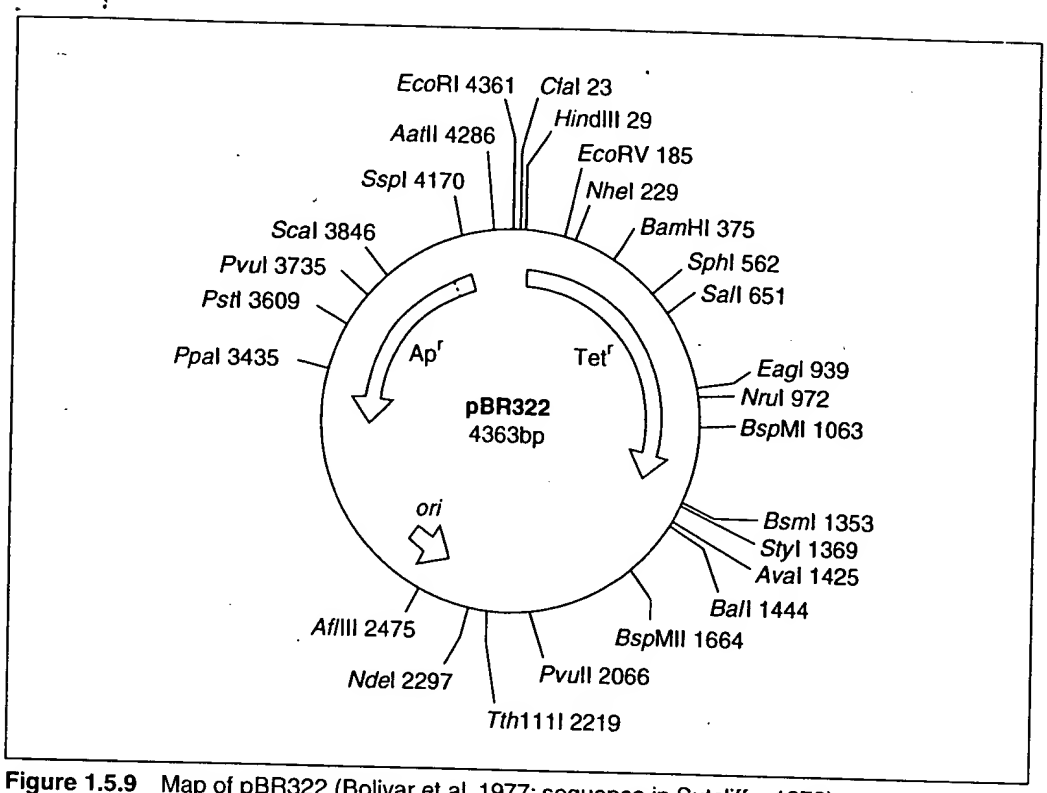

Figure 1.5.9 Map of pBR322 (Bolivar et al. 1977; sequence in Sutcliffe, 1978).

pBR322 is one of the classic cloning vectors from which many other vectors are derived. It contains an amplifiable pMB1 replicator and genes encoding resistance to ampicillin and tetracycline. Insertion of DNA into a restriction site in either drug-resistance gene usually inactivates it and allows colonies bearing plasmids with such insertions to be identified by their inability to grow on medium with that antibiotic (see Fig. 1.5.9; Bolivar et al., 1977; sequence in Sutcliffe, 1978).

pUC19 belongs to a family of plasmid vectors that contains a polylinker inserted within the alpha region of the *lacZ* gene. The polylinkers are the same as those used in the m13mp series (Fig. 1.14.2). pUC19 and pUC18 have the same polylinker but in opposite orientations. Under appropriate conditions (see *UNIT 1.4* for a description), colonies that bear plasmids containing a fragment inserted into the polylinker form white colonies instead of blue ones. These pMB1-derived plasmids (see Fig. 1.5.2) maintain a very high-copy-number (1000 to 3000 per genome). Wild-type and recombinant plasmids confer ampicillin resistance and can be amplified with chloramphenicol (Norrander et al., 1983). In addition wild-type plasmids confer a LacZ$^+$ phenotype to appropriate cells (e.g., JM101 cells, *UNIT 1.4*).

pBluescript is a commonly used phagemid cloning vector that contains a polylinker inserted into the alpha region of the *lacZ* gene and T3 and T7 promoter sequences flanking the cloning sites. The f1 (+) filamentous phage origin of replication in pBluescript SK$^+$ allows for the recovery of the sense strand of the *lacZ* gene as ssDNA; the pBluescript SK(−) vector (Fig. 1.5.10) with the f1 origin in the opposite orientation, f1(−), facilitates recovery of the other strand. The position of the polylinker in the alpha region of the *lacZ* gene allows for identification of inserts based on a blue/white color screen under the appropriate conditions. The T3 and T7 promoters are recognized by bacteriophage RNA polymerases. Transcription from these promoters reads into the polylinker from either side..RNA transcripts of any DNA cloned into the polylinker can thus be produced by run-off transcription in vitro.

pWE15 is an example of a cosmid vector used for cloning DNA fragments ~35 to 45 kb (see Fig. 1.5.3; Wahl et al., 1987). The *cos* sites allow the DNA to be cut and packaged into phage heads by the appropriate lambda proteins. There is a single unique *Bam*HI cloning site flanked by T3 and T7 promoter sequences. These promoters are particularly useful for production of labeled RNA probes corresponding to the ends of the insert DNA, and these can be used to identify overlapping cosmids for chromosome walking and construction of cosmid contigs. *Not*1 sites flanking the cloning site can

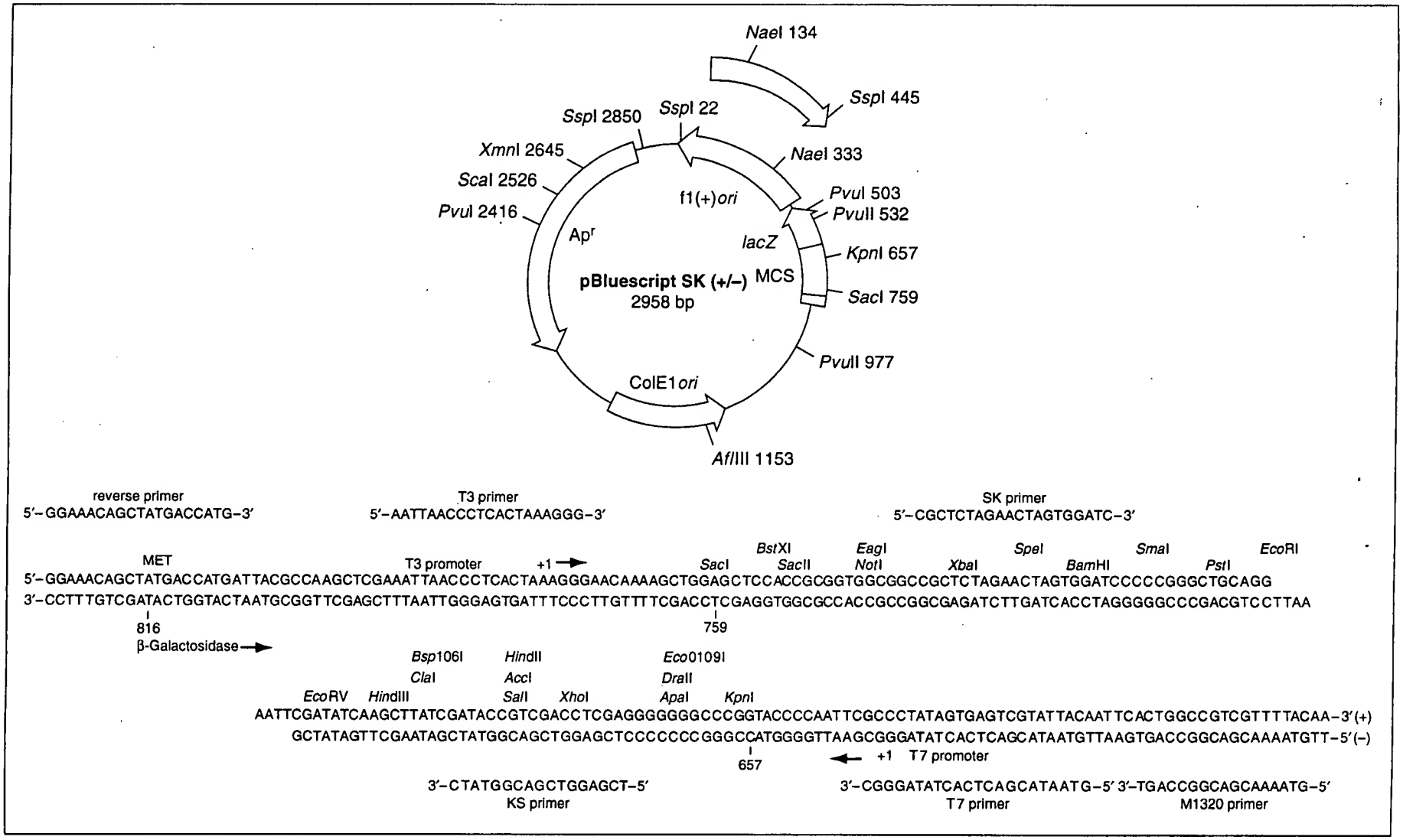

Figure 1.5.10 Map of pBluescript SK (+/–).

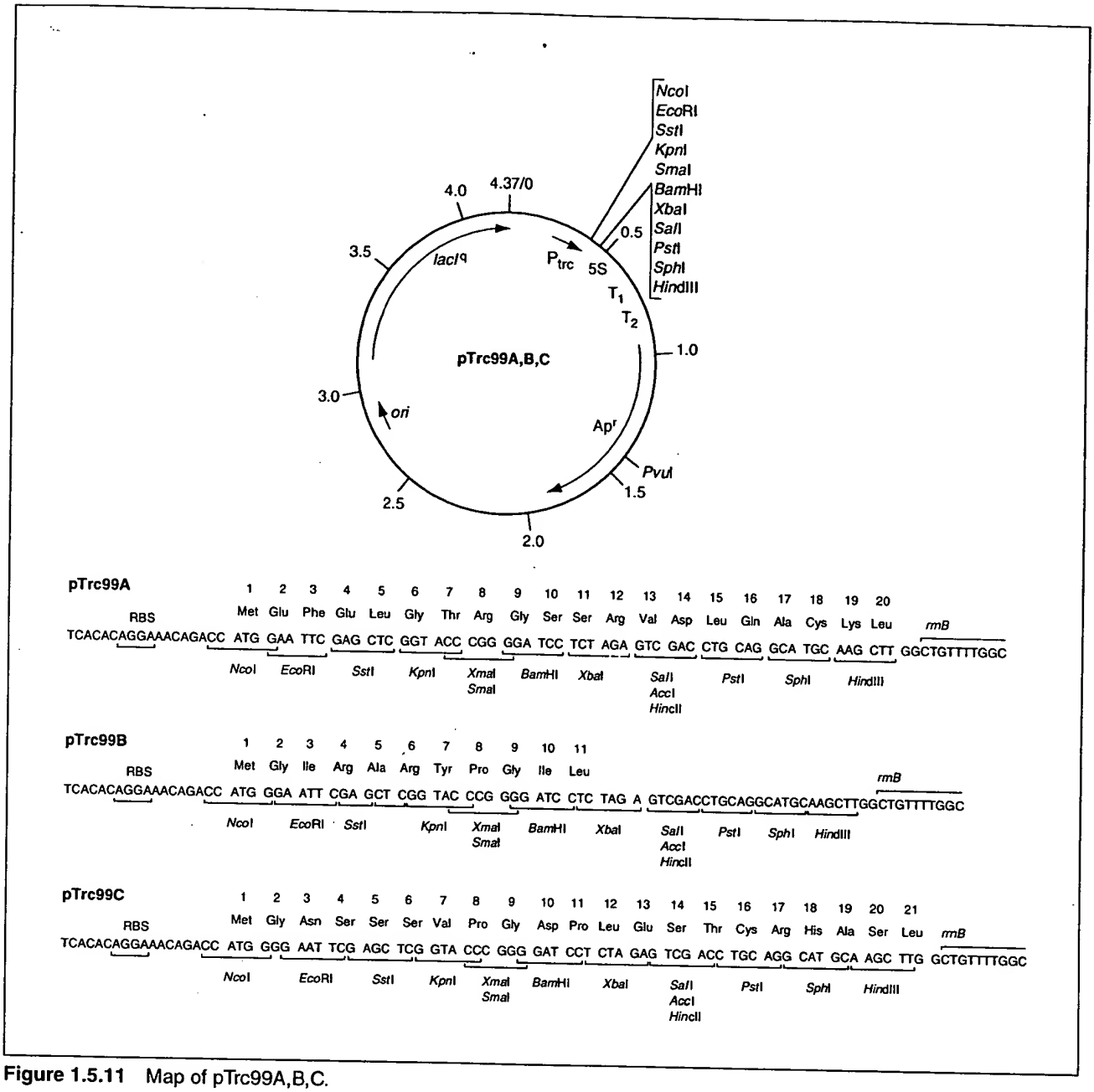

Figure 1.5.11 Map of pTrc99A,B,C.

potentially be used to excise an intact insert fragment from the vector. The ColE1-derived *ori* and ampicillin resistance gene allow for replication and selection in bacteria. The SV40 promoter (included in SV40 *ori*) which drives the neomycin phosphotransferase gene enables selection in eukaryotic cells.

pBeloBAC11 is an example of the family of bacterial-artificial-chromosome vectors based on the low-copy-number F factor replicator (see Fig. 1.5.4; Shizuya et. al., 1992). BAC vectors are used for cloning large DNA fragments (100 to 500 kb) in *E. coli* and are used commonly in genome-mapping strategies. *oriS*, *repE*, *parA*, *parB*, and *parC* genes are the essential genes that compose the F factor replicator. *oriS* and *repE* genes are required for unidirectional replication of the plasmid, and *parABC* loci stably maintain the copy number at one to two per *E. coli* genome. There are two unique cloning sites (*Hin*dIII and *Bam*HI) inserted into the *lacZ* alpha region. Other useful features of the cloning region are (1) T7 and SP6 promoter sequences flanking the cloning sites, (2) *Not*I restriction sites flanking the cloning sites for potential excision of the insert, and (3) and presence of the *loxP* and *cosN* sites that can be cleaved by specific enzymes. The ends generated by cleavage at *loxP* or *cosN* can be used as fixed reference points in building an

ordered restriction map by end labeling and partial restriction digestion.

pEGFP-1 is a selectable vector for monitoring promoter activity in mammalian cells via fluorescence of a green fluorescent protein (GFP) derivative (Clontech; see Fig. 1.5.5; Yang et al., 1996). The vector contains a neomycin resistance gene downstream of the SV40 early promoter for selection of stably transformed mammalian cells. It has a polylinker located upstream of the EGFP gene, so that the function of promoter sequences introduced into the polylinker can be assessed based on EGFP activity. The EGFP gene is modified from wild-type GFP to ensure expression in mammalian cells, it has silent base mutations that correspond to human codon-usage preferences, and sequences flanking the coding region have been converted to a Kozak consensus translation initiation signal. The vector backbone contains an f1 *ori* for production of ssDNA, a pUC-derived *ori* for propagation in *E. coli* and a kanamycin resistance gene for selection in bacteria.

A series of yeast shuttle vectors (pRS304, 305, and 306) has been created to facilitate manipulation of DNA in *Saccharomyces cerevisiae* (see Fig. 1.5.6; Sikorski and Hieter, 1989). These vectors have a backbone derived from pBLUESCRIPT into which the features necessary for replication and maintenance in yeast have been introduced. The members of this series of plasmids differ only in the yeast selectable marker incorporated; pRS303 carries the *HIS3* marker that complements a non-reverting *his3* chromosomal mutation in specific yeast strains. These plasmids contain an autonomously replicating sequence as well as a centromere sequence, *CEN6*, that ensures stable maintenance in yeast cells.

pcDNA3.1 is a selectable cloning and expression vector for use in mammalian cells. The features of this vector include a neomycin resistance gene driven by the SV40 early promoter (contained within the SV40 *ori*) and terminated by an SV40 polyadenylation signal for selection in mammalian cells (see Fig. 1.5.7). In addition, due to the inclusion of the SV40 *ori*, the vector can replicate as an episome in cells expressing the SV40 large T antigen. The polylinker cloning site is located downstream of strong cytomegalovirus enhancer-promoter sequences and upstream of the bovine growth hormone gene termination signals for high-level expression of protein-coding sequences cloned into this vector. This vector also contains some of the more standard features of other plasmid vectors, including a ColE1 replicator for propagation in *E. coli*, the ampicillin resistance gene for selection in *E. coli*, the f1 *ori* for production of ssDNA, and the T7 promoter sequence for in vitro transcription of DNA inserted into the polylinker.

pRR54 is an example of a broad-host-range mobilizable plasmid vector. This vector contains replicator and stablilization sequences derived from the natural RK2 broad-host-range plasmid (see Fig. 1.5.8; Roberts et al., 1990). *oriV* is the vegetative origin of replication, *trfA* encodes *trans*-acting functions necessary for replication, and *par* encodes a locus that enhances stability of the plasmid. This plasmid can be mated into diverse gram-negative species as long as the appropriate mobilization machinery is provided in *trans* because it contains the origin of conjugal transfer, *oriT*. The plasmid carries the β-lactamase gene, allowing for ampicillin/carbenecillin selection of plasmid containing bacteria.

The pTrc series of plasmid expression vectors facilitates regulated expression of genes in *E. coli*. These vectors carry the strong hybrid trp/lac promoter, the *lacZ* ribosome-binding site (RBS), the MCS of pUC18 that allows insertion in three reading frames, and the *rrnB* transcription terminators (see polylinker sequences given below the vector diagram in Fig. 1.5.11). These vectors are equally useful for expression of unfused proteins (resulting from insertion in the *Nco*I site) or for expression of fusion proteins (using one of the cloning sites in the correct translational frame). The presence of the *lacI*^{q} allele on the plasmid ensures complete repression of the hybrid trp/lac promoter during cloning and growth in any host strain (see Amann et al., 1988, for further details).

LITERATURE CITED

Amann, E., Ochs, B., and Abel, K.-J. 1988. Tightly regulated *tac* promoter vectors useful for the expression of unfused and fused proteins in *Escherichia coli*. *Gene* 69:301-315.

Bolivar, F., Rodriguez, R.L., Greene, P.J., Betlach, M.C., Heynecker, H.L, and Boyer, H.W. 1977. Construction of useful cloning vectors. *Gene* 2:95-113.

Chang, A.C.Y. and Cohen, S.N. 1978. Construction and characterization of amplifiable mulitcopy DNA cloning vehicles derived from the P15A cryptic miniplasmid. *J. Bacteriol.* 134:1141-1156.

Gerhart, E., Wagner, H., and Simmons, R.W. 1994. Antisense RNA control in bacteria, phages, and plasmids. *Annu. Rev. Microbiol.* 48:713-742.

Jackson, D.A., Symons, R.M., and Berg, P. 1972. Biochemical method for inserting new genetic information into DNA of Simian Virus 40 circular DNA molecules containing lambda phage genes and the galactose operon of *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.* 69:2904-2909.

Kahn, M., Kolter, R., Thomas, C., Figurski, D., Meyer, R., Remaut, E., and Helinski, D.R. 1979. Plasmid cloning vehicles derived from plasmids ColE1, F, R6K, RK2. *Methods Enzymol.* 68:268-280.

Kim, U.J., Shizuya, H., de Jong, P.J., Birren, B., and Simon, M.I. 1992. Stable propagation of cosmid-sized human DNA inserts in an F-factor based vector. *NAR* 20:1083-1085.

Norrander, J., Kempe, T., and Messing, J. 1983. Construction of improved M13 vectors using oligonucleotide-directed mutagenesis. *Gene* 26:101-106.

Roberts, R.C., Burioni, R., and Helinski, D.R. 1990. Genetic characterization of the stabilizing functions of a region of broad-host-range plasmid RK2. *J. Bacteriol.* 172:6204-6216.

Shizuya, H., Birren, B., Kim U.J., Mancin, V., Slepak, T., Tachiiri, Y., and Simon, M.I. 1992. A bacterial system for cloning large human DNA fragments. *Proc. Natl. Acad. Sci. U.S.A.* 89:8794-8797.

Sikorski, R.S. and Hieter, P.I. 1989. A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in *Saccharomyces cerevisiae*. *Genetics* 122:19-27.

Stoker, N.G., Fairweather, N., and Spratt, B.G. 1982. Versatile low-copy-number plasmid vectors for cloning in *Escherichia coli*. *Gene* 18:335-341.

Sutcliffe, J.G. 1978. Complete nucleotide sequence of the *Escherichia coli* plasmid pBR322. *Cold Spring Harbor Symp. Quant. Biol.* 43:77-90.

Uhlin, B.E., Schweickart, V., and Clark, A.J. 1983. New runaway-replication-plasmid cloning vectors and suppression of runaway replication by novobiocin. *Gene* 22:255-265.

Wahl, G.M., Lewis, K.A., Ruiz, J.C., Rothenberg, B., Zhao, J., and Evans, G.A. 1987. Cosmid vectors for rapid genomic walking, restriction mapping, and gene transfer. *Proc. Natl. Acad. Sci. U.S.A.* 84:2160-2164.

Yang, T., Cheng, L., and Kain, S.R. 1996. Optimized codon usage and chromophore mutations provide enhanced sensitivity with the green fluorescent protein. *Nucl. Acids Res.* 24:4592-4593.

Contributed by Rhonda Feinbaum
Massachusetts General Hospital
Boston, Massachusetts